Topology and Predictability of El Niño and La Niña Networks

Anastasios A. Tsonis and Kyle L. Swanson

Department of Mathematical Sciences, Atmospheric Sciences Group, University of Wisconsin–Milwaukee, Milwaukee, Wisconsin 53201, USA

(Received 15 January 2008; published 5 June 2008)

We construct the networks of the surface temperature field for El Niño and for La Niña years and investigate their structure. We find that the El Niño network possesses significantly fewer links and lower clustering coefficient and characteristic path length than the La Niña network, which indicates that the former network is less communicative and less stable than the latter. We conjecture that because of this, predictability of temperature should decrease during El Niño years. Here we verify that indeed during El Niño years predictability is lower compared to La Niña years.

DOI: 10.1103/PhysRevLett.100.228502 PACS numbers: 92.10.am, 05.40.−a, 89.60.−k, 89.75.−k

The study of networks has recently exploded into a major research tool in many areas of sciences. The discovery of ''small-world'' and scale-free networks has led to many new insights about complex systems in many areas of science [1–6]. Lately, networks have also been applied to climate [7–9]. Constructing a climate network assumes that climate is represented by a grid of oscillators, each one of them representing a dynamical system varying in some complex way. What we are interested in is the collective behavior of these interacting dynamical systems and the structure of the resulting network. These applications of ideas from graph theory to climate have resulted in the discovery of novel insights about the role of atmospheric teleconnections in climate [8] and how the synchronization of major climate modes may lead to significant climate shifts [9].

Here we present a new application where we construct the networks for the global surface temperature field separately for El Niño years and for La Niña years. More specifically we consider the global National Center for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis surface temperature data set [10]. The data are arranged on a grid with a resolution of 5° latitude × 5° longitude. This results in 72 points in the east-west direction and 37 points in the north-south direction for a total of $n = 2664$ points. These 2664 points will be assumed to be the nodes of the network. For each grid point monthly values from 1950–2005 are available. From the monthly values we produced anomaly values (actual value minus the climatological average for each month). Even though the leading order effect of the annual cycle is removed by producing anomaly values, some of it is still present as the amplitude of the anomalies is greater in the winter than in the summer. For this reason, in order to avoid spurious high values of correlations, only the values for November–March in each year were considered. Thus, for each grid point we have a time series of 330 anomaly values. These 330 values are then separated into El Niño values if they fall during an El Niño event, into La Niña values if the fall during a La Niña event, and a residual if they fall in between. These values are then concatenated to produce for each node a time series for El Niño and a time series for La Niña. Here an El Niño event is defined when the normalized El Niño/Southern Oscillation (ENSO) index is greater than 1 and a La Niña event is defined when the normalized ENSO index is less than −1. These limits represent the ± standard deviation from the mean. It turns out that the resulted sample size in El Niño and La Niña years is very similar (58 for El Niño and 59 for La Niña). In order to define the links between the nodes for either network, the correlation coefficient at lag zero (r) between the time series of all possible pairs of nodes [$n(n-1)/2 = 3,547,116$ pairs] is estimated. It follows that the estimation of the correlation coefficient between any two time series is based on a sample size of about 60. Note that since the values are monthly anomalies there is very little autocorrelation in the time series. A pair is considered as connected if the absolute value of their cross correlation $|r| \geq 0.5$. This criterion is based on parametric and nonparametric significance tests. According to the t test with $N = 60$, a value of $r = 0.5$ is statistically significant above the 99% level [11]. In addition, randomization experiments, where the values of the time series of one node in a pair are scrambled and then are correlated to the unscrambled values of the time series of the other node, indicate that a value of $r = 0.5$ will not arise by chance. The use of the correlation coefficient to define links in networks is not new. Correlation coefficients have been used to successfully derive the topology of gene expression networks [12–15] and to study financial markets [16].

The architecture of the resulted network is presented in Fig. 1, which shows the area weighted number of total links at each geographic location for El Niño (top) and La Niña (bottom). More accurately, it shows the fraction of the total global area to which a point is connected. This is an appropriate way to show the architecture of a network when it is a continuous network defined on a sphere.

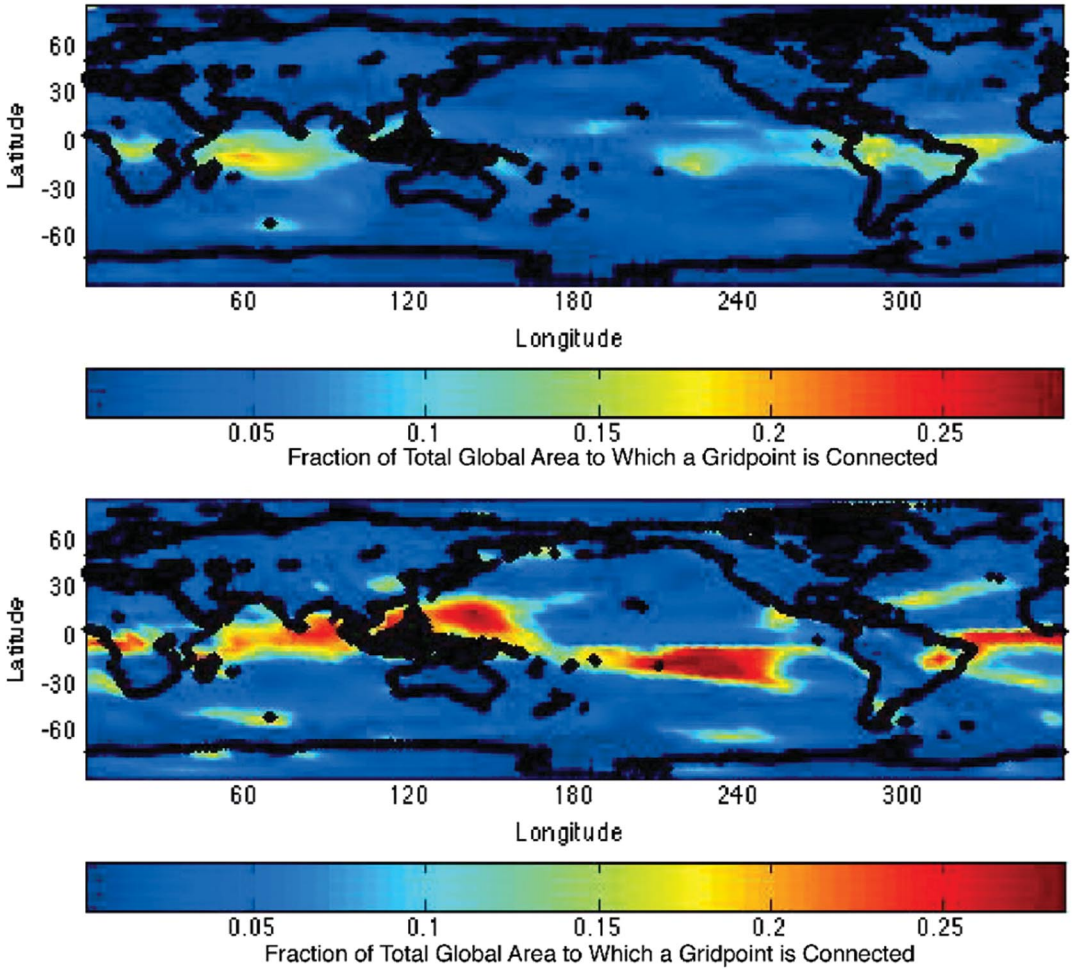

FIG. 1 (color). Total number of links at each geographic location (grid point) as delineated by the fraction of the total global area that a point is connected to. The top panel corresponds to El Niño years and the bottom panel to La Niña years. Both networks are characterized by a presence of supernodes (nodes with many links) albeit less pronounced in the El Niño network.

Thus, if a node i is connected to N other nodes at λ_N latitudes, its area weighted connectivity $\tilde{C}_i$ is defined as

$$\tilde{C}_i = \sum_{j=1}^{N} \cos\lambda_j \Big/ \sum_{\text{over all } \lambda \text{ and } \varphi} \cos\lambda, \qquad (1)$$

where φ is the longitude. In the above expression the denominator is the area of Earth's surface and the numerator is the area of that surface a node is connected to. Both networks are characterized by a presence of supernodes (nodes with many links) albeit less pronounced in the El Niño network. Networks with supernodes are often scale-free networks with a power law degree distribution (the degree distribution p_k gives the probability that a node in the network is connected to k other nodes). Indeed, their degree distribution is approximately a power law (Fig. 2). The conclusion from Fig. 1 is that during El Niño the network loses many of its links. Moreover, as Fig. 3 indicates, it appears that this loss applies to all spatial scales. Figure 3 shows the frequency distribution of links as a function of link length. Clearly, El Niño exhibits fewer links than La Niña at all scales. We then find that this loss of links is associated with a significant decrease in the network's clustering coefficient and characteristic path length, two measures that characterize the topology of the network. The clustering coefficient for each node is estimated by considering the closest neighborhood of the node given by the corresponding $\tilde{C}_i$ and finding the fraction of this neighborhood that is connected. The average C_i over all nodes provides the clustering coefficient of the network C [7,8,17]. The characteristic path length is the average number of links in the shortest path between two nodes [17]. We find that the clustering coefficient and characteristic path length in the El Niño network are 40% and 60% smaller, respectively. This indicates that the El Niño network is less communicative than the La Niña network, which in turn makes the El Niño network more prone to failure (more unstable) [17–19].

We conjecture that this tendency during El Niño events will be associated with loss of predictability. Figure 4

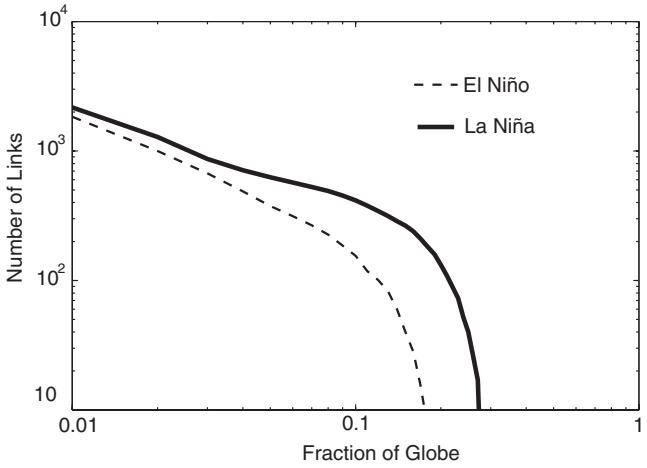

FIG. 2. Degree distribution for the La Niña network (solid line) and the El Niño network (broken line). For most of the scales involved, the distribution is a power law.

shows the average variance of the surface temperature for the tropics (20°S–20°N) from persistence for 10 day lead forecasts, which measure how the trajectory diverges in the near future at a given location. To maintain consistency with the networks shown in Fig. 1, all forecast and verification quantities are averaged over a 5° latitude × 5° longitude grid as well. Surface temperature variance from persistence is accumulated for each day during the cold season and then averaged for each month to extract the climate signal in that variance associated with ENSO variability. To understand the magnitude of this signal, it must be noted that forecasting surface temperature in the tropics is an extremely challenging problem. Day-to-day variation in tropical surface temperature is primarily determined by the interaction between the surface and moist convective motions in the atmosphere. These motions impact the transfer of energy between the surface and the overlying atmosphere via radiative, sensible heat, and latent heat transfers, which are essentially random and inject variability into the system at all time scales [20]. Because of this randomness and the difficulty in resolving the complicated underlying physical processes, computer models have little or no demonstrable skill in forecasting tropical surface temperatures.

Given the above, it is expected that any predictability signal will be subtle. Nevertheless, we find that the monthly averaged variance of surface temperature from persistence, shown in Fig. 4, is greater during El Niño events. This slope (indicated by the solid straight line) according to the t test differs from zero with a significance $p < 0.01$. Additional support is provided by the F test, where the null hypothesis is equality of the variances for El Niño and La Niña. If we consider only El Niño (ENSO index $> +1$) and La Niña events (ENSO index < -1) (shown on the right and on the left of the vertical broken lines, respectively), we find that the null hypothesis is rejected as on the average the monthly averaged tropical surface temperature variance from persistence during El Niño conditions is about 20% higher than during La Niña conditions ($F = 1.19$, $p < 0.1$). This change in the time evolution and hence predictability of the tropical surface temperature is consistent with the network properties above.

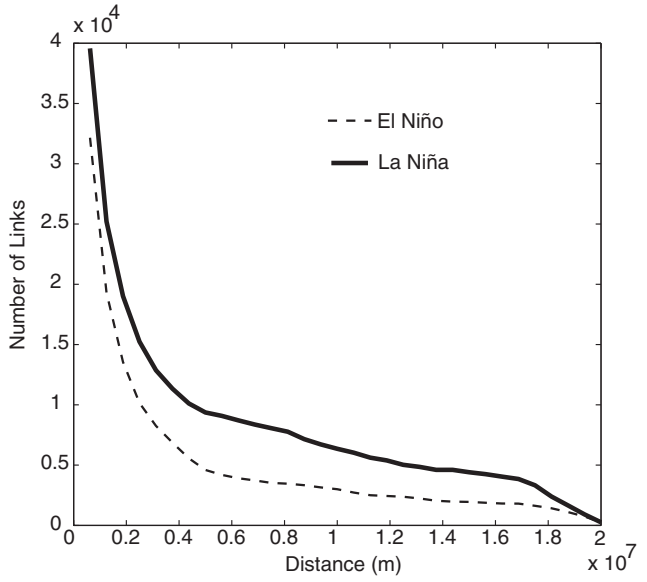

FIG. 3. The relative frequency distribution of links as a function of link length. Clearly during El Niño events the number of links at all space scales is reduced.

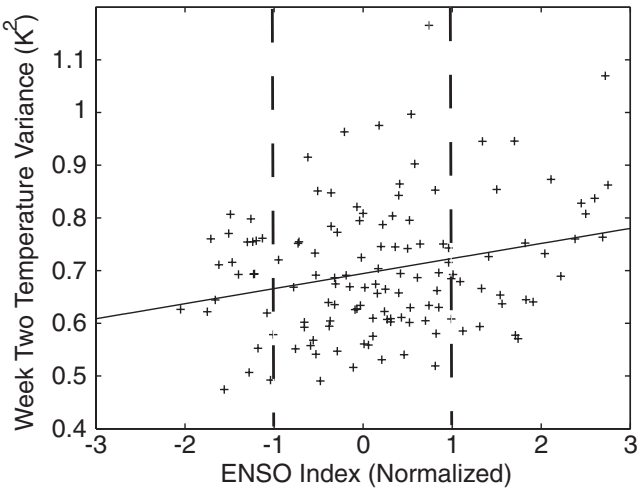

FIG. 4. Monthly averaged variance of the surface temperature for the tropics (20°S–20°N) from persistence for 10 day lead forecasts at a given location. Here data in the period 1979–2006 are used. Prior to that tropical daily temperature fields are less reliable due to poor satellite coverage. As explained in the text, this figure suggests that for El Niño events the variance is increased compared to La Niña events indicating loss of predictability during El Niño events.

El Niño and global temperature are known to have an intricate relationship, in which one is forcing the other in a rather nonlinear way [21,22]. We now find that this relationship includes an effect on the predictability of the temperature field, a fact not previously known. Our results are important as they indicate that certain dynamics may be responsible for the loss in predictability during El Niño. We speculate that El Niño, being a departure from normal, introduces fluctuations over all space scales. These fluctuations act as extra noise into the system, which limit predictability. Recognizing and understanding these effects may result in improvements of temperature forecasting during El Niño events.

The authors are supported by NSF Grant No. ATM-0438612.

[1] R. Albert, H. Jeong, and A.-L. Barabasi, Nature (London) **401**, 130 (1999).
[2] H. Jeong, S. Mason, A.-L. Barabasi, and Z. N. Oltvai, Nature (London) **411**, 41 (2001).
[3] H. Jeong, B. Tombor, R. Albert, A. N. Oltvai, and A.-L. Barabasi, Nature (London) **407**, 651 (2000).
[4] F. Liljeros, C. Edling, L. N. Amaral, H. E. Stanley, and Y. Aberg, Nature (London) **411**, 907 (2001).
[5] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[6] J.-P. Bouchaud and M. Mezard, Physica (Amsterdam) **282A**, 536 (2000).
[7] A. A. Tsonis, K. L. Swanson, and P. J. Roebber, Bull. Am. Meteorol. Soc. **87**, 585 (2006).
[8] A. A. Tsonis, K. L. Swanson, and G. Wang, J. Clim. (to be published).
[9] A. A. Tsonis, K. L. Swanson, and S. Kravtsov, Geophys. Res. Lett. **34**, L13705 (2007).
[10] R. Kistler *et al.*, Bull. Am. Meteorol. Soc. **82**, 247 (2001).
[11] The choice of $r = 0.5$, while it guarantees statistical significance, is somewhat arbitrary. We find that while other values might affect the connectivity structure of the network, the effect of different correlation thresholds is negligible on the conclusions reached in this study.
[12] I. J. Farkas, H. Jeong, T. Vicsek, A.-L. Barabási, and Z. N. Oltvai, Physica (Amsterdam) **318A**, 601 (2003).
[13] D. E. Featherstone and K. Broadie, BioEssays **24**, 267 (2002).
[14] A. de la Fuente, P. Brazhnik, and P. Mendes, Trends Genet. **18**, 395 (2002).
[15] H. Agrawal, Phys. Rev. Lett. **89**, 268702 (2002).
[16] R. N. Mantegna, Eur. Phys. J. B **11**, 193 (1999).
[17] R. Albert and A.-L. Barabasi, Rev. Mod. Phys. **74**, 47 (2002).
[18] R. Albert, H. Jeong, and A.-L. Barabasi, Nature (London) **406**, 378 (2000).
[19] A.-L. Barabasi and E. Bonabeau, Sci. Am. **288**, No. 5, 60 (2003).
[20] K. A. Emanuel, *Atmospheric Convection* (Oxford University, New York, 1994).
[21] A. A. Tsonis, A. G. Hunt, and J. B. Elsner, Meteorol. Atmos. Phys. **84**, 229 (2003).
[22] A. A. Tsonis, J. B. Elsner, A. G. Hunt, and T. H. Jagger, Geophys. Res. Lett. **32**, L09701 (2005).